

Online Checkpointing with Improved Worst-Case Guarantees

Karl Bringmann* Benjamin Doerr Adrian Neumann
Jakub Sliacan

May 1, 2013

Abstract

In the online checkpointing problem, the task is to continuously maintain a set of k checkpoints that allow to rewind an ongoing computation faster than by a full restart. The only operation allowed is to replace an old checkpoint by the current state. Our aim are checkpoint placement strategies that minimize rewinding cost, i.e., such that at all times T when requested to rewind to some time $t \leq T$ the number of computation steps that need to be redone to get to t from a checkpoint before t is as small as possible. In particular, we want that the closest checkpoint earlier than t is not further away from t than q_k times the ideal distance $T/(k+1)$, where q_k is a small constant.

Improving over earlier work showing $1 + 1/k \leq q_k \leq 2$, we show that q_k can be chosen asymptotically less than 2. We present algorithms with asymptotic discrepancy $q_k \leq 1.59 + o(1)$ valid for all k and $q_k \leq \ln(4) + o(1) \leq 1.39 + o(1)$ valid for k being a power of two. Experiments indicate the uniform bound $p_k \leq 1.7$ for all k . For small k , we show how to use a linear programming approach to compute good checkpointing algorithms. This gives discrepancies of less than 1.55 for all $k < 60$.

We prove the first lower bound that is asymptotically more than one, namely $q_k \geq 1.30 - o(1)$. We also show that optimal algorithms (yielding the infimum discrepancy) exist for all k .

1 Introduction

Checkpointing means storing selected intermediate states of a long sequence of computations. This allows reverting the system to an arbitrary previous state much faster, since only the computations from the preceding checkpoint have to be redone. Checkpointing is one of the fundamental techniques in computer science. Classic results date back to the seventies [3], more recent topics are checkpointing in distributed [4], sensor network [7], or cloud [10] architectures.

*Karl Bringmann is a recipient of the *Google Europe Fellowship in Randomized Algorithms*, and this research is supported in part by this Google Fellowship.

Checkpointing usually involves a careful trade-off between the speed-up of reversions to previous states and the costs incurred by setting checkpoints (time, memory). Much of the classic literature (see [5] and the references therein) studies checkpointing with the focus of gaining fault tolerance against immediately detectable faults. Consequently, only reversions to the most recent checkpoint are needed. However, setting a checkpoint can be highly time consuming, because the whole system state has to be copied to secondary memory. In such scenarios, the central question is how often to set a checkpoint such that the expected time spent on setting checkpoints and redoing computations from the last checkpoint is minimized (under a stochastic failure model and further, possibly time-dependent [9], assumptions on the cost of setting a checkpoint).

In this work, we will regard a checkpointing problem of a different nature. If not fault-tolerance of the system is the aim of checkpointing, then often the checkpoints can be kept in main memory. Applications of this type arise in data compression [2] and numerics [6, 8]. In such scenarios, the cost of setting a checkpoint is small compared to the cost of the regular computation. Consequently, the memory used by the stored checkpoints is the bottleneck.

The first to provide an abstract framework independent of a particular application in mind were Ahlroth, Pottonen and Schumacher [1]. They do not make assumptions on which reversion to previous states will be requested, but simply investigate how checkpoints can be set in an online fashion such that at all times their distribution is balanced over the total computation history.

They assume that the system is able to store up to k checkpoints (plus a free checkpoint at time 0). At any point in time, a previous checkpoint may be discarded and replaced by the current system state as new checkpoint. Costs incurred by such a change are ignored. However, as it turns out, good checkpointing algorithms do not set checkpoints very often. For all algorithms discussed in the remainder of this paper, each checkpoint is changed only $O(\log T)$ times up to time T .

The max-ratio discrepancy measure. Each set of checkpoints, together with the current state and the state at time 0, partitions the time from the process start to the current time T into $k + 1$ disjoint intervals. Clearly, without further problem-specific information, an ideal set of checkpoints would lead to all these intervals having identical length. Of course, this is not possible at all points in time due to the restriction that new checkpoints can only be set on the current time. As discrepancy measure for a checkpointing algorithm, Ahlroth et al. mainly regard the *maximum gap ratio*, that is, the maximum ratio of the longest interval vs. the shortest interval (ignoring the last interval, which can be arbitrarily small), over all current times T . They show that there is a simple algorithm achieving a discrepancy of two: Start with all checkpoints placed evenly, e.g., at times $1, \dots, k$. At an even time T , remove one of the checkpoints at an odd time and place it at T . This will lead to all checkpoints being at the even times $2, 4, \dots, 2k$ when $T = 2k$ is reached. Since these checkpoints form a scaled copy of the initial ones, we can continue in this fashion forever. It is easy

to see that at all times, the intervals formed by neighboring checkpoints have at most two different lengths, the larger being twice the smaller in case that not all lengths are equal. This shows the discrepancy of two.

It seems tempting to believe that one can do better, but, in fact, not much improvement is possible for general k as shown by the lower bound of $2^{1-1/\lceil(k+1)/2\rceil} = 2(1 - o(1))$. For small values of k , namely $k = 2, 3, 4$, and 5 , better upper bounds of approximately $1.414, 1.618, 1.755$, and 1.755 , respectively, were shown.

The maximum distance discrepancy measure. In this work, we shall regard a different, and, as we find, more natural discrepancy measure. Recall that the actual cost of reverting to a particular state is basically the cost of redoing the computation from the preceding checkpoint to the desired point in time. Adopting a worst-case view on the time to revert to, our aim is to keep the length of the longest interval small (at all times). Note that with time progressing, the interval lengths necessarily grow. Hence a fair point of comparison is the length $T/(k+1)$ of a longest interval in the (at time T) optimal partition of the time frame into equal length intervals. For this reason, we say that a checkpointing algorithm (using k checkpoints) has *maximum distance discrepancy* (or simply discrepancy) q if it places the checkpoints in such a way that at all times T , the longest interval has length at most $qT/(k+1)$. We denote by $q^*(k)$ the infimum discrepancy among all checkpointing algorithms using k checkpoints.

This maximum distance discrepancy measure was suggested in [1]. There it was remarked that an upper bound of β for the gap-ratio discrepancy implies an upper bound of $\beta(1 + \frac{1}{k})$ for the maximum distance discrepancy. Furthermore, for all k an upper bound of 2 and a lower bound of $1 + \frac{1}{k}$ is shown for $q^*(k)$. For $k = 2, 3, 4$, and 5 , stronger upper bounds of $1.785, 1.789, 1.624$, and 1.565 , respectively, were shown.

Our results. In this work, we show that the optimal discrepancy $q^*(k)$ is asymptotically bounded away from both one and two by a constant. We present algorithms that achieve a discrepancy of $1.59 + O(1/k)$ for all k (Theorem 2), and a discrepancy of $\ln(4) + o(1) \leq 1.39 + o(1)$ for k being any power of two (Theorem 3). For small values of k , and this might be an interesting case in applications with memory-consuming states, we show superior bounds by suggesting a class of checkpointing algorithms and optimizing their parameters via a combination of exhaustive search and linear programming (Table 1). Experiments suggest $q^*(k) \leq 1.7$ for all k (Sect. 6). We complement these constructive results by a lower bound for $q^*(k)$ of $2 - \ln(2) - O(1/k) \geq 1.3 - O(1/k)$ (Theorem 6). We round off this work with a natural, but seemingly nontrivial result: We show that for each k there is indeed a checkpointing algorithm having discrepancy $q^*(k)$ (Theorem 4). In other words, the infimum in the definition of $q^*(k)$ can be replaced by a minimum.

2 Notation and Preliminaries

In the checkpointing problem with k checkpoints, we consider a long running computation during which we can choose to save the state at the current time T in a checkpoint, or delete a previously placed one. We assume that our storage can hold at most k checkpoints simultaneously, and that there are implicit checkpoints at time $t = 0$ and the current time. We disregard any costs for placing or maintaining checkpoints. Consequently, we may assume that we only delete a previous checkpoint when a new one is placed.

An *algorithm* for checkpoint placement can be described by two infinite sequences. First, the time points where new checkpoints are placed, i.e., a non-decreasing infinite sequence of reals $t_1 \leq t_2 \leq \dots$ such that $\lim_{i \rightarrow \infty} t_i = \infty$, and second, a rule that describes which old checkpoints to delete when a new one is installed, that is, an injective function $d : [k+1.. \infty) \rightarrow \mathbb{N}$ satisfying $d_i < i$ for all $i \geq k+1$.

The algorithm A described by (t, d) will start with t_1, \dots, t_k as initial checkpoints and then for each $i \geq k+1$, at time t_i remove the checkpoint at t_{d_i} and set a new checkpoint at the current time t_i . We call the act of removing a checkpoint and placing a new one a *step* of A . Note that there is little point in setting the first k checkpoints to zero, so to make the following discrepancy measure meaningful, we shall always require that $t_k > 0$.

We call the set of checkpoints that exist at time T *active*. The active checkpoints, together with the two implicit checkpoints at times 0 and T , define a sequence of $k+1$ interval lengths $\mathcal{L}_T = (\ell_0, \dots, \ell_k)$. The *discrepancy* $q(A, T)$ of an algorithm A at time $T \geq t_k$ is a measure of how long the maximal interval is, normalized to be one if all intervals have the same length. It is calculated as

$$q(A, T) := (k+1)\bar{\ell}_T/T,$$

where $\bar{\ell}_T = \|\mathcal{L}_T\|_\infty$ denotes the length of the longest interval. We also use the term *discrepancy* when we refer to the scaled length of a single interval.

The discrepancy $\text{Perf}(A)$ of an algorithm A then is the supremum over the discrepancy over all times T , i.e.,

$$\text{Perf}(A) := \sup_{T \geq t_k} q(A, T).$$

Hence the discrepancy of an algorithm would be 1, if it kept its checkpoints evenly distributed at all times. Denote the infimum discrepancy of a checkpointing algorithm using k checkpoints by

$$q^*(k) := \inf_A \text{Perf}(A),$$

where A runs over all algorithms using k checkpoints. We will see in Sect. 7 that algorithms achieving this discrepancy actually exist.

Note that we allow checkpointing algorithms to set checkpoints at continuous time points. One can convert any such algorithm to an algorithm with integral

checkpoints by rounding all checkpointing times t_i down. This does not increase the discrepancy since $\lfloor t_i \rfloor - \lfloor t_{i-1} \rfloor \leq t_i - t_{i-1} + 1$, but with discrete time there are at most $\lfloor t_i \rfloor - \lfloor t_{i-1} \rfloor - 1$ steps to recompute in this interval.

In the definition of the discrepancy, the supremum is never attained at some T with $t_i < T < t_{i+1}$ for any i , as shown in the following lemma.

Lemma 1. *In the definition of the discrepancy it suffices to consider times $T = t_i$ for all $i \geq k$, i.e., we have*

$$\text{Perf}(A) = \sup_{i \geq k} q(A, t_i).$$

Proof. Consider a time T with $t_i < T < t_{i+1}$ for any $i \geq k$. We show that

$$q(A, T) \leq \max\{q(A, t_i), q(A, t_{i+1})\}.$$

Denote the active checkpoints at time T by x_1, \dots, x_k . Note that $x_k = t_i$, since t_i was the last time we set a checkpoint. Consider the interval $[x_k, T]$. Its discrepancy is exactly

$$(k+1) \frac{T - x_k}{T} \leq (k+1) \frac{t_{i+1} - x_k}{t_{i+1}} \leq q(A, t_{i+1}).$$

Any other interval at time T is of the form $[x_{j-1}, x_j]$ for some $1 \leq j \leq k$ (where we set $x_0 := 0$), whose discrepancy is

$$(k+1) \frac{x_j - x_{j-1}}{T} \leq (k+1) \frac{x_j - x_{j-1}}{t_i} \leq q(A, t_i).$$

Together, this proves the claim. \square

To bound the discrepancy of an algorithm we need to bound the largest of the $q(A, t_i)$ over all $i \geq k$. For this purpose, it suffices to look at the two newly created intervals at time t_i for each i , as made explicit by the following lemma.

Lemma 2. *Let $i > k$ and let ℓ_1, ℓ_2 be the lengths of the two newly created intervals at time t_i due to the removal and the insertion of a checkpoint. Then*

$$\max\{q(A, t_{i-1}), q(A, t_i)\} = \max\{q(A, t_{i-1}), (k+1)\ell_1/t_i, (k+1)\ell_2/t_i\}.$$

Proof. If ℓ_1 or ℓ_2 is the longest interval at time t_i the claim holds. Any other interval existed already at time t_{i-1} and had a larger discrepancy at this time, as we divide by the current time to compute the discrepancy. Thus, if any other interval is the longest at time t_i , then we have $q(A, t_{i-1}) \geq q(A, t_i)$ and the claim holds again. \square

Often, it will be useful to use a different notation for the checkpoint that is removed in step i . Instead of the global index d , one can also use the index $p : [k+1.. \infty) \rightarrow [1..k]$ among the active checkpoints, i.e.,

$$p_i = d_i - |\{j \in [i-1] \mid d_j < d_i\}|.$$

We call an algorithm $A = (t, p)$ *cyclic*, if the p_i are periodic with some period n , i.e., $p_i = p_{i+n}$ for all i , and after n steps A has transformed the intervals to a scaled version of themselves, that is, $\mathcal{L}_{t_{k+jn}} = \gamma^j \mathcal{L}_{t_k}$ for some $\gamma > 1$ and all $j \in \mathbb{N}$. We call γ the *scaling factor*. For a cyclic algorithm A , it suffices to fix the *pattern* of removals $P = (p_{k+1}, \dots, p_{k+n})$ and the checkpoint positions $t_1, \dots, t_k, t_{k+1}, \dots, t_{k+n}$. Since our discrepancy notion is invariant under scaling, we can assume without loss of generality that $t_k = 1$ (and hence $t_{k+n} = \gamma$).

Since cyclic algorithms transform the starting position to a scaled copy of itself, it is easy to see that their discrepancy is given by the maximum over the discrepancies during one period, i.e., for cyclic algorithms A with period n we have

$$\text{Perf}(A) = \max_{k < i \leq k+n} q(A, t_i).$$

This makes this class of algorithms easy to analyze.

3 Introductory Example – A Simple Bound for $k = 3$

For the case of $k = 3$ there is a very simple algorithm, SIMPLE, with a discrepancy of $4/\phi^2 \approx 1.53$, where $\phi = (\sqrt{5} + 1)/2$ is the golden ratio. Because the algorithm is so simple, we use it to familiarize ourselves with the notation we introduced in Sect. 2. The algorithm is cyclic with a pattern of length one. We prove the following theorem.

Theorem 1. *For $k = 3$ there is a cyclic algorithm SIMPLE with period length one and*

$$\text{Perf}(\text{SIMPLE}) = \frac{4}{\phi^2}.$$

Proof. We fix the pattern to be $P = (1)$, that is, algorithm SIMPLE always removes the oldest checkpoint. For this simple pattern it is easy to calculate the discrepancy depending on the scaling factor γ . Since the intervals need to be a scaled copy of themselves after just one step and we can fix $t_3 = 1$, we know immediately that

$$t_1 = \frac{1}{\gamma^2}, \quad t_2 = \frac{1}{\gamma}, \quad t_3 = 1, \quad t_4 = \gamma,$$

and hence the discrepancy is determined by

$$4 \cdot \max \left\{ \frac{t_1 - 0}{t_3}, \frac{t_2 - t_1}{t_3}, \frac{t_3 - t_2}{t_3} \right\} = 4 \cdot \max \left\{ \frac{1}{\gamma^2}, \frac{\gamma - 1}{\gamma^2}, \frac{\gamma - 1}{\gamma} \right\}.$$

Since $\gamma > 1$, the second term is always smaller than the third and can be ignored. As $1/\gamma^2$ is decreasing and $(\gamma - 1)/\gamma$ is increasing, the maximum is minimal when they are equal. Simple calculation shows this to be the case at $\gamma = \phi$.

Hence for $k = 3$ the algorithm with pattern (1) and checkpoint positions $t_1 = 1/\phi^2$, $t_2 = 1/\phi$, $t_3 = 1$, and $t_4 = \phi$ has discrepancy $4/\phi^2 \approx 1.53$. \square

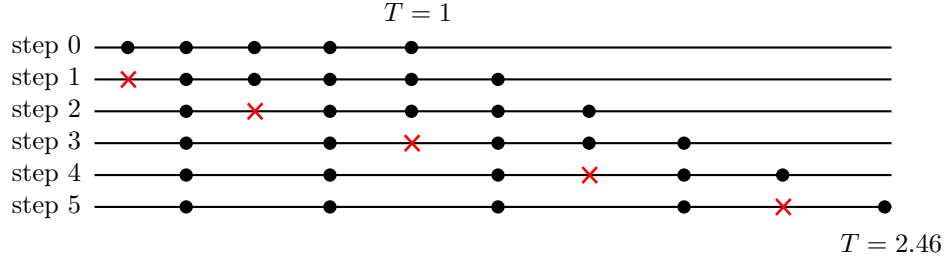


Figure 1: One period of the algorithm LINEAR from Sect. 4 for $k = 5$. After one period all intervals are scaled by the same factor.

The experiments in Sect. 6 indicate that for $k = 3$ this is optimal among all cyclic algorithms with a period of length at most 6.

4 A Simple Upper Bound for Large k

In this section we present an algorithm, LINEAR, with a discrepancy of roughly 1.59 for large k . This improves upon the asymptotic bound of 2 from [1]. Moreover, LINEAR is easily implemented for all k .

Like the algorithm SIMPLE of the previous section, the algorithm LINEAR is cyclic. It has a simple pattern of length k . The pattern is just $(1, \dots, k)$, that is, at the i -th step of a period LINEAR deletes the i -th active checkpoint. Overall, during one period LINEAR removes all checkpoints at times t_i with odd index i , as shown in Fig. 1.

This removal pattern is identical to the one of POWERS-OF-TWO algorithm from [1]. However, that algorithm starts with a uniform checkpoint distribution where removing any checkpoint doubles the maximum interval. This leads to an asymptotic discrepancy of two. In contrast, LINEAR places checkpoints on a polynomial. For $i \in [1, 2k]$ we set $t_i = (i/k)^\alpha$, where α is a constant. In the analysis we optimize the choice of α and set $\alpha := 1.302$. For this algorithm we show the following theorem.

Theorem 2. *Algorithm LINEAR has a discrepancy of at most*

$$\text{Perf}(\text{LINEAR}) \leq 1.586 + O(k^{-1}).$$

Experiments show that the discrepancy of algorithm LINEAR is close to the bound of 1.586 even for moderate sizes of k . Comparisons using the optimization method from Sect. 6 indicate that for the pattern $(1, \dots, k-1)$ of algorithm LINEAR, different checkpoint placements can yield only improvements of about 4.5% for large k . Experimental results are summarized in Fig. 4.

Proof. As algorithm LINEAR is cyclic, we can again compute the discrepancy from the $2k$ checkpoint positions and the pattern,

$$\text{Perf}(\text{LINEAR}) = \max_{k < i \leq 2k} (k+1)\bar{\ell}_{t_i}/t_i,$$

where $\bar{\ell}_{t_i}$ is the length of the longest interval at time t_i . By Lemma 2 it suffices to consider newly created intervals at times t_{k+1}, \dots, t_{2k} . Note that at time t_i we create the intervals $[t_{i-1}, t_i]$ (from insertion of a checkpoint at t_i) and $[t_{2(i-k)-2}, t_{2(i-k)}]$ (from deletion of the checkpoint at $t_{2(i-k)-1}$). The discrepancy of the new interval by insertion is, for $k < i \leq 2k$,

$$(k+1) \frac{t_i - t_{i-1}}{t_i} = (k+1) \frac{i^\alpha - (i-1)^\alpha}{i^\alpha} \leq (k+1) \frac{(k+1)^\alpha - k^\alpha}{(k+1)^\alpha}.$$

Using $(x+1)^c - x^c \leq c(x+1)^{c-1}$ for any $x \geq 0$ and $c \geq 1$, this simplifies to

$$\leq (k+1) \frac{\alpha(k+1)^{\alpha-1}}{(k+1)^\alpha} = \alpha,$$

for any constant $\alpha \geq 1$.

For the new interval from deleting the checkpoint at $t_{2(i-k)-1}$ we get a discrepancy of

$$\begin{aligned} (k+1) \frac{t_{2(i-k)} - t_{2(i-k)-2}}{t_i} &= (k+1) \frac{(2(i-k))^\alpha - (2(i-k)-2)^\alpha}{i^\alpha} \\ &\leq (k+1) 2^\alpha \frac{\alpha(i-k)^{\alpha-1}}{i^\alpha}, \end{aligned}$$

where we used again $(x+1)^c - x^c \leq c(x+1)^{c-1}$. An easy computation shows that $(i-k)^{\alpha-1}/i^\alpha$ is maximized at $i = \alpha k$ over $k < i \leq 2k$. Hence, we can upper bound this discrepancy by

$$\leq \left(1 + \frac{1}{k}\right) 2^\alpha \frac{\alpha(\alpha-1)^{\alpha-1}}{\alpha^\alpha} = 2^\alpha \left(1 - \frac{1}{\alpha}\right)^{\alpha-1} + O(k^{-1}).$$

We optimize the latter term numerically and obtain for $\alpha = 1.302$ an upper bound of

$$1.586 + O(k^{-1}).$$

Note that this bound is larger than the bound $\alpha = 1.302$ from the new intervals from insertion. Hence, overall we get the desired upper bound. \square

5 An Improved Upper Bound for Large k

In this section we present the algorithm **BINARY** that yields a discrepancy of roughly $\ln(4) \approx 1.39$ for large k . Compared to the algorithm **LINEAR** from the last section, **BINARY** has a considerably better discrepancy at the price of a more involved analysis, and it only works for k being a power of two.

Theorem 3. *For $k \geq 8$ being any power of 2, the algorithm **BINARY** has discrepancy*

$$\text{Perf}(\text{BINARY}) \leq \ln(4) + \frac{0.05}{\lg(k/4)} + O\left(\frac{1}{k}\right).$$

Here and in the remainder of this paper, let ‘lg’ denote the binary and ‘ln’ the natural logarithm. Note that the term $O(1/k)$ quickly tends to 0, whereas the $\Theta(1/\lg(k/4))$ term is small due to the constant 0.05. Hence, this discrepancy is close to $\ln(4)$ already for moderate k . Also note that $\ln(4)$ is by less than 0.1 larger than our lower bound from Sect. 8, leaving room for less than a 6% improvement over the upper bound for algorithm BINARY for large k . We verified experimentally that algorithm BINARY yields very good bounds already for relatively small k . The results are summarized in Fig. 5.

5.1 The Algorithm BINARY

The initial checkpoints t_1, \dots, t_k satisfy the equation

$$t_i = \alpha t_{i/2} \quad (1)$$

for each even $1 \leq i \leq k$ and some $\alpha = \alpha(k) \geq 2$. Precisely, we set

$$\alpha := 2^{1 + \frac{\lg(\sqrt{2}/\ln 4)}{\lg(k/4)}} \approx 2^{1 + \frac{0.029}{\lg(k/4)}}.$$

However, the usefulness of this expression becomes clear only in the analysis of the algorithm.

During one period we delete all odd checkpoints t_1, t_3, \dots, t_{k-1} and insert the new checkpoints

$$t_{k+i} := \alpha t_{k/2+i}, \quad (2)$$

for $1 \leq i \leq k/2$. Then after one period we end up with the checkpoints

$$= \alpha \cdot \begin{pmatrix} t_2, t_4, \dots, t_{k-2}, t_k, & t_{k+1}, t_{k+2}, \dots, t_{k+k/2} \\ (t_1, t_2, \dots, t_{k/2-1}, t_{k/2}, & t_{k/2+1}, t_{k/2+2}, \dots, t_{k/2+k/2}) \end{pmatrix} = \alpha(t_1, t_2, \dots, t_k),$$

which proves cyclicity. Note that (1) and (2) allow us to compute all t_i from the values $t_{k/2+1}, \dots, t_k$, however, we still have some freedom to choose the latter values. Without loss of generality we can set $t_k := 1$, then $t_{k/2} = \alpha^{-1}$. In between these two values, we interpolate $\lg t_i$ linearly, i.e., we set for $i \in (k/2, k]$

$$t_i := \alpha^{2^{i/k-2}}, \quad (3)$$

completing the definition of the t_i . Note that this equation also works for $i = k$ and $i = k/2$.

There is one more freedom we have with this algorithm, namely in which order we delete all odd checkpoints during one period, i.e., we need to fix the pattern of removals. In iteration $1 \leq i \leq k/2$ we insert the checkpoint t_{k+i} and remove the checkpoint $t_{d(i+k)}$, defined as follows. For $m \in \mathbb{N} = \mathbb{N}_{\geq 1}$ let $2^{e(m)}$ be the largest power of 2 that divides m . We define $S: \mathbb{N} \rightarrow \mathbb{N}, S(m) := m/2^{e(m)}$. Note that $S(m)$ is an odd integer. Using this definition, we set

$$d(k+i) := S\left(i + \frac{k}{2}\right), \quad (4)$$

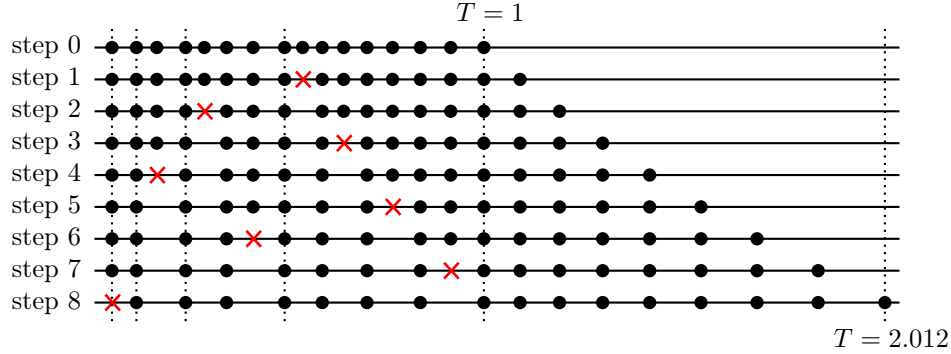


Figure 2: One period of the algorithm BINARY for $k = 16$. Note that, recursively, checkpoints are removed twice as often from the right half of the initial setting (at steps i where $i \bmod 2 = 1$) as from the second quarter.

finishing the definition of the algorithm BINARY. If we write this down as a pattern, then we have $p_i = 1 + k/(2^{1+e(i)})$ for $1 \leq i < k/2$ and $p_{k/2} = 1$. For intuition as to the behavior of this pattern, see the example in Fig. 2. The following lemma implies that the deletion behavior of BINARY is indeed well-defined, meaning that during one period we delete all odd checkpoints t_1, t_3, \dots, t_{k-1} (and no point is deleted twice).

Lemma 3. *The function S induces a bijection between $\{k/2 < i \leq k\}$ and $\{1 \leq i \leq k \mid i \text{ is odd}\}$.*

Proof. Let $A := \{k/2 < i \leq k\}$ and $B := \{1 \leq i \leq k \mid i \text{ is odd}\}$. Since $S(m) \leq m$ and $S(m)$ is odd for all $m \in \mathbb{N}$, we have $S(A) \subseteq B$. Moreover, A and B are of the same size. We present an inverse function to finish the proof. Let $x \in B$. Note that there is a unique number $y \in \mathbb{N}$ such that $x2^y \in A$, since A is a range between two consecutive powers of 2 and $x \leq k$. Setting $S^{-1}(x) = x2^y$ we have found the inverse. \square

5.2 Discrepancy Analysis

We now bound the largest discrepancy encountered during one period, i.e.,

$$\text{Perf}(\text{BINARY}) = \max_{1 \leq i \leq k/2} q(\text{BINARY}, t_{i+k}) = (k+1) \max_{1 \leq i \leq k/2} \bar{\ell}_{t_{i+k}}/t_{i+k}.$$

We first compute the maximum and later multiply with the factor $k+1$. By Lemma 2, we only have to consider intervals newly created by insertion and deletion at any step.

Intervals from Insertion: We first compute the discrepancy of the interval newly added at time t_{i+k} , $1 \leq i \leq k/2$. Its length is $t_{i+k} - t_{i+k-1}$, so its

discrepancy (without the factor $k + 1$) is

$$\begin{aligned}\frac{t_{i+k} - t_{i+k-1}}{t_{i+k}} &= 1 - \frac{t_{i+k-1}}{t_{i+k}} \\ &= 1 - \frac{t_{i+k/2-1}}{t_{i+k/2}} \\ &\stackrel{(3)}{=} 1 - \alpha^{-2/k},\end{aligned}$$

where the second equality holds because of (2) if $i > 1$ or (1) if $i = 1$.

Using $e^x \geq 1 + x$ for $x \in \mathbb{R}$ yields a bound on the discrepancy of

$$\frac{t_{i+k} - t_{i+k-1}}{t_{i+k}} \leq \ln(\alpha) \frac{2}{k} = \ln(\alpha^2).$$

Deleting t_1 : We show similar bounds for the intervals we get from deleting an old checkpoint. We first analyze the deletion of t_1 —this case is different from the general one, since t_1 has no predecessor. Note that t_1 is deleted at time $t_{3k/2}$. The deletion of t_1 creates the interval $[0, t_2]$. This interval has discrepancy

$$\frac{t_2}{t_{3k/2}} \stackrel{(2),(1)}{=} \frac{\alpha t_1}{\alpha t_k} \stackrel{(1)}{=} \alpha^{-\lg k} \leq 1/k,$$

since we choose $\alpha \geq 2$. Hence, this discrepancy is dominated by the one we get from newly inserted intervals.

Other Intervals from Deletion: It remains to analyze the discrepancy of the intervals we get from deletion in the general case, i.e., at some time t_{i+k} , $1 \leq i < k/2$. At this time we delete checkpoint $d(i+k)$, so we create the interval $[t_{d(i+k)-1}, t_{d(i+k)+1}]$ of discrepancy

$$q_i := \frac{t_{d(i+k)+1} - t_{d(i+k)-1}}{t_{i+k}} \stackrel{(2),(4)}{=} \frac{t_{S(i+k/2)+1} - t_{S(i+k/2)-1}}{\alpha t_{i+k/2}}.$$

Let $h := e(i + k/2)$, so that 2^h is the largest power of 2 dividing $i + k/2$, and $2^h S(i + k/2) = i + k/2$. Then $t_{S(i+k/2)+1} = \alpha^{-h} t_{i+k/2+2^h}$ by (1), and a similar statement holds for $t_{S(i+k/2)-1}$, yielding

$$q_i = \alpha^{-1-h} \frac{t_{i+k/2+2^h} - t_{i+k/2-2^h}}{t_{i+k/2}}.$$

Using (3) we get $t_{i+k/2} = \alpha^{2i/k-1}$. Comparing this with the respective terms for $t_{i+k/2+2^h}$ and $t_{i+k/2-2^h}$ yields

$$\begin{aligned}q_i &= \alpha^{-1-h} \left(\alpha^{2^{h+1}/k} - \alpha^{-2^{h+1}/k} \right) \\ &= \alpha^{-1-h} \cdot 2 \sinh(\ln(\alpha^2) 2^h/k).\end{aligned}$$

By elementary means one can show that the function $f(x) = x^{-A} \sinh(Bx)$, $A \geq 1, B > 0$, is convex on $\mathbb{R}_{\geq 0}$. Since convex functions have their maxima at the boundaries of their domain, and since by above equation q_i can be expressed using $f(2^h)$ (for $A = \lg \alpha$ and $B = \ln(\alpha^2)/k$), we see that q_i is maximal at (one of) the boundaries of h . Recall that we treated $i = k/2$ separately, and observe that the largest power of 2 dividing $i + k/2$, $1 \leq i < k/2$ is at most $k/4$. Hence, we have $0 \leq 2^h \leq k/4$ and

$$q_i \leq \max \{2\alpha^{-1} \sinh(\ln(\alpha^2)/k), 2\alpha^{-1}(k/4)^{-\lg \alpha} \sinh(\ln(\alpha)/2)\}.$$

We simplify using $\alpha \geq 2$ and $\sinh(x) = x + O(x^2)$ to get

$$q_i \leq \max \{ \ln(\alpha^2)/k + O(1/k^2), (k/4)^{-\lg \alpha} \sinh(\ln(\alpha)/2) \}. \quad (5)$$

The first term is already of the desired form. For the second one, note that setting $\alpha = 2$ we would get a discrepancy of $4 \sinh(\ln(2)/2)/k = \sqrt{2}/k$. We get a better bound by choosing

$$\alpha := 2^{1 + \frac{c}{\lg(k/4)}},$$

with $c := \lg(\sqrt{2}/\ln(4)) \approx 0.029$. Then the second bound on q_i from above becomes

$$(k/4)^{-\lg \alpha} \sinh(\ln(\alpha)/2) = \frac{4}{k} 2^{-c} \sinh\left(\frac{\ln(2)}{2} \left(1 + \frac{c}{\lg(k/4)}\right)\right).$$

The particular choice of c allows to bound the derivative of $\sinh((1+x)\ln(2)/2)$ for $x \in [0, c]$ from above by

$$\frac{\ln(2)}{2} \cosh((1+c)\ln(2)/2) < 0.39.$$

Hence, we can upper bound

$$\sinh\left(\frac{\ln(2)}{2} \left(1 + \frac{c}{\lg(k/4)}\right)\right) \leq \sinh(\ln(2)/2) + \frac{0.39c}{\lg(k/4)}.$$

Thus, in total the second bound on q_i from inequality (5) becomes

$$(k/4)^{-\lg \alpha} \sinh(\ln(\alpha)/2) \leq \frac{4}{k} 2^{-c} \sinh(\ln(2)/2) + \frac{4 \cdot 2^{-c} \cdot 0.39c}{k \lg(k/4)}.$$

Since $c = \lg(\sqrt{2}/\ln(4)) = \lg(4 \sinh(\ln(2)/2)/\ln(4))$, this becomes

$$\leq \ln(4)/k + 0.044/(k \lg(k/4)).$$

Overall discrepancy: In total, we can bound the discrepancy $q := \text{Perf}(\text{BINARY})$ of our algorithm (now including the factor of $k+1$) by

$$q \leq (k+1) \max \{ \ln(\alpha^2)/k + O(1/k^2), \ln(4)/k + 0.044/(k \lg(k/4)) \}.$$

Using $(k+1)/k = 1 + O(1/k)$ and

$$\ln(\alpha^2) = \ln(4) \left(1 + \frac{c}{\lg(k/4)} \right) \leq \ln(4) + \frac{0.040}{\lg(k/4)},$$

this bound can be simplified to

$$q \leq \max\{\ln(4) + 0.040/\lg(k/4) + O(1/k), \ln(4) + 0.044/\lg(k/4) + O(1/k)\},$$

which proves Theorem 3.

6 Upper Bounds via Combinatorial Optimization

In this section we show how to find upper bounds on the optimal discrepancy $q^*(k)$ for fixed k . We do so by constructing cyclic algorithms using exhaustive enumeration of all short patterns in the case of very small k or randomized local search on the patterns for larger k , combined with linear programming to optimize the checkpoint positions. This yields good algorithms as summarized in Table 1. In the following we describe our algorithmic approach.

Finding Checkpoint Positions: First we describe how to find a nearly optimal cyclic algorithm given a pattern P and a scaling factor γ , i.e., how to optimize the checkpoint positions. To do so, we construct a linear program that is feasible if a cyclic algorithm with discrepancy λ and scaling factor γ exists. We use three kinds of constraints: We fix the ordering of the checkpoints, enforce that the i -th active checkpoint after one period is a factor γ larger than the i -th initial checkpoint, and upper bound the discrepancy of each interval during the period by λ . We then use binary search to optimize λ .

Lemma 4. *For a fixed pattern P of length n and scaling factor γ , let $q^* = \inf_A \text{Perf}(A)$ be the optimal discrepancy among algorithms A using P and γ . Then finding an algorithm with discrepancy at most $q^* + \epsilon$ reduces to solving $O(\log \epsilon^{-1})$ linear feasibility problems with $O(nk)$ inequalities and $k + n$ variables.*

Proof. For a fixed pattern and scaling factor, we can tune the discrepancy of the algorithm by cleverly choosing the time points when to remove an old checkpoint and place a new one. By solving a linear feasibility problem we can check whether a cyclic algorithm with scaling factor γ and pattern P exists that guarantees a discrepancy of at most λ . We can then optimize over λ to find an approximately optimal algorithm.

We construct a linear program with the $k + n$ time points (t_1, \dots, t_{k+n}) as variables (where we can set $t_k = 1$ without loss of generality). It uses three kinds of constraints. The first kind is of the form

$$t_i \leq t_{i+1},$$

for all $i \in [1, k + n)$. These constraints are satisfied if the checkpoint positions have the correct ordering, i.e. checkpoints with larger index are placed at later times.

The second kind of constraints enforces the scaling factor. Since the pattern is fixed, we can compute at all steps which checkpoints are active. For $i \in [1, k]$ and $j \in [0, n]$, let τ_i^j be the variable of the i -th active checkpoint in step j and let τ_0^j be 0 for all j . It is easy to see that the algorithm has a scaling factor of γ if the i -th active checkpoint in the last step is larger by a factor of γ than in the first step. We encode this as constraints of the form

$$\tau_i^n = \gamma \tau_i^0.$$

Lastly we encode an upper bound of λ for the discrepancy. Since the discrepancy of a cyclic algorithm is given by

$$\max_{k < i \leq k+n} (k+1) \bar{\ell}_{t_i} / t_i,$$

and each $\bar{\ell}_{t_i}$ can be expressed by a maximum over k terms, we can encode a discrepancy guarantee of λ with nk constraints of the form

$$\tau_{i+1}^j - \tau_i^j \leq \lambda \tau_k^j / (k+1),$$

for all $i \in [0, k)$ and $j \in [0, n]$.

A feasible solution of these constraints fixes the checkpoint positions and hence, together with the pattern P , provides an algorithm with discrepancy at most λ . Using a simple binary search over $\lambda \in [1, 2]$ we can find an approximately optimal algorithm for this value of γ and the pattern P . \square

Finding Scaling Factors: Next we show how to find scaling factors γ for which algorithms with good discrepancy exist. We first show an upper bound for γ .

Lemma 5. *A cyclic algorithm with k checkpoints, discrepancy $\lambda < k$, and a period length of n can have scaling factor at most*

$$\gamma \leq \left(\frac{1}{1 - \lambda/(k+1)} \right)^n.$$

Proof. Consider any checkpointing algorithm $A = (t, d)$ with k checkpoints and discrepancy λ . At any time t_i , $i \geq k$, the largest interval has length $\bar{\ell}_{t_i} \geq t_i - t_{i-1}$, as there is no checkpoint in the time interval $[t_{i-1}, t_i]$. Hence, we have

$$(k+1) \frac{t_i - t_{i-1}}{t_i} \leq \lambda.$$

Rearranging, this yields

$$t_i \leq \frac{1}{1 - \lambda/(k+1)} t_{i-1}.$$

Iterating this n times, we get

$$t_{k+n} \leq \left(\frac{1}{1 - \lambda/(k+1)} \right)^n t_k.$$

k	3	4	5	6	7	8	9	10	15	20	30	50	100
Discr.	1.529	1.541	1.472	1.498	1.499	1.499	1.488	1.492	1.466	1.457	1.466	1.481	1.484

Table 1: Upper bounds for different k . For $k < 8$ all patterns up to length k were tried. For $k = 8$ all patterns up to length 7 were tried. For larger k , patterns were found via randomized local search.

Hence, for any cyclic algorithm (with discrepancy λ , k checkpoints, and a period length of n) we get the desired bound on the scaling factor $\gamma = t_{k+n}/t_k$. \square

Since algorithms with discrepancy 2 are known [1], we can restrict our attention to $\lambda \leq 2$. Hence, for any given pattern length n , Lemma 5 yields an upper bound on γ , while a trivial lower bound is given by $\gamma > 1$. Now, for any given pattern P we optimize over γ using a linear search with a small step size over the possible values for γ . For each tested γ , we optimize over the checkpoint positions using the linear programming approach described above.

Finding Patterns: For small k and n , we can exhaustively enumerate all k^n removal patterns of period length n . Some patterns can be discarded as they obviously cannot lead to a good algorithm or are equivalent to some other pattern: No pattern that never removes the first checkpoint can be cyclic. Furthermore, patterns are equivalent under cyclic shifts, so we can assume without loss of generality that all patterns end with removing the first checkpoint. Lastly, it never makes sense to remove the currently last checkpoint. Hence, for k checkpoints there are at most $(k-1)^{n-1}$ interesting patterns of length n . This finishes the description of our combinatorial optimization approach.

Results: We ran experiments that try patterns up to length k for $k \in [3, 7]$. For $k = 8$ we stopped the search after examining patterns of length 7. For larger k we used a randomized local search to find good patterns. The upper bounds we found are summarized in Table 1, and for $k \leq 8$ the removal patterns and time points when to place new checkpoints can be found in Fig. 3. Note that for $k = 3$ this procedure re-discovers the golden ratio algorithm of Sect. 3.

Note that we can combine the results presented in Table 1 with the algorithm LINEAR (Theorem 2 and Fig. 4) to read off a global upper bound of $q^*(k) \leq 1.7$ for the optimal discrepancy for *any* k .

For a fixed pattern the method is efficient enough to find good checkpoint positions for much larger k . For $k \leq 1000$ we experimentally compared the algorithm LINEAR of Sect. 4 with algorithms found for its pattern $(1, \dots, k-1)$. The experiments show that for $k = 1000$ LINEAR is within 4.5% of the optimized bounds. For the algorithm BINARY of Sect. 5, this comparison is even more favorable. For $k = 1024$ the algorithm places its checkpoints so well that the optimization procedure improves discrepancy only by 1.9%. The results are summarized in Fig. 4 and Fig. 5.

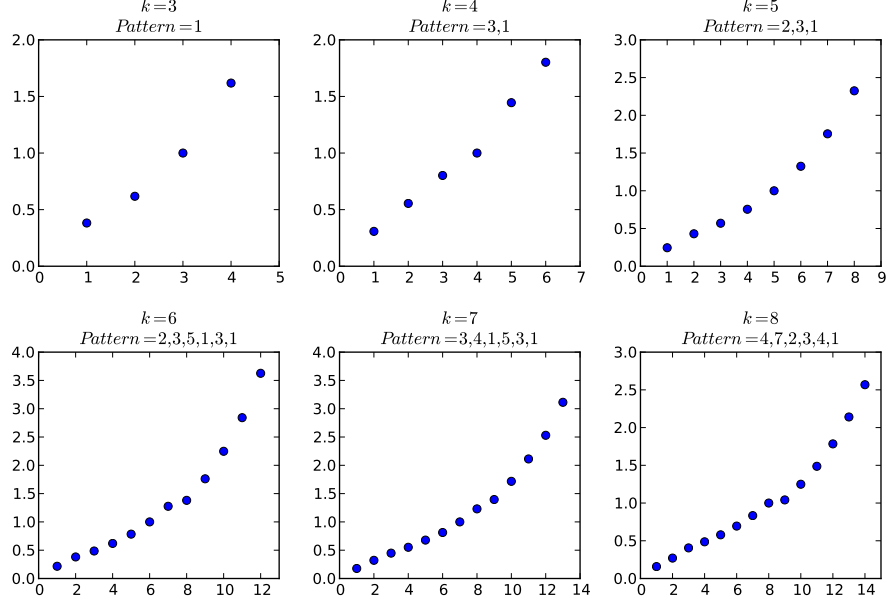


Figure 3: Time points where the i -th checkpoint is placed to achieve the bounds of Table 1. Time is on the y -Axis, iteration is on the x -Axis.

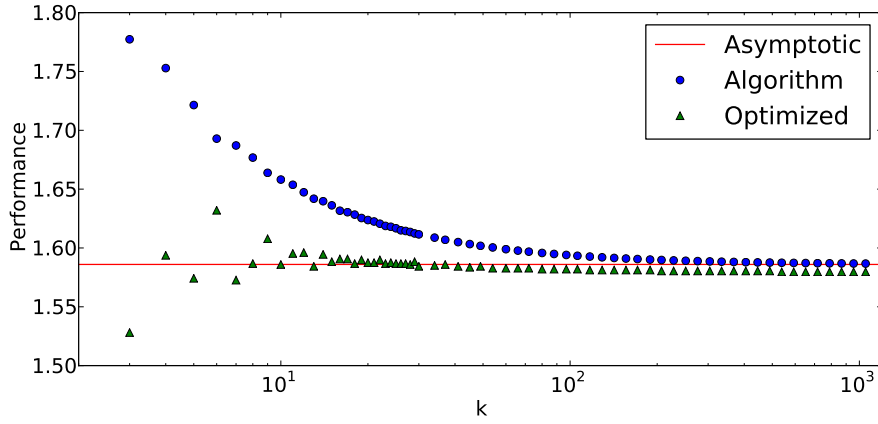


Figure 4: The discrepancy of algorithm LINEAR from Sect. 4 for different values of k compared with the upper bounds for its pattern found via the combinatorial method from Sect. 6. For large k LINEAR is about 4.5% worse.

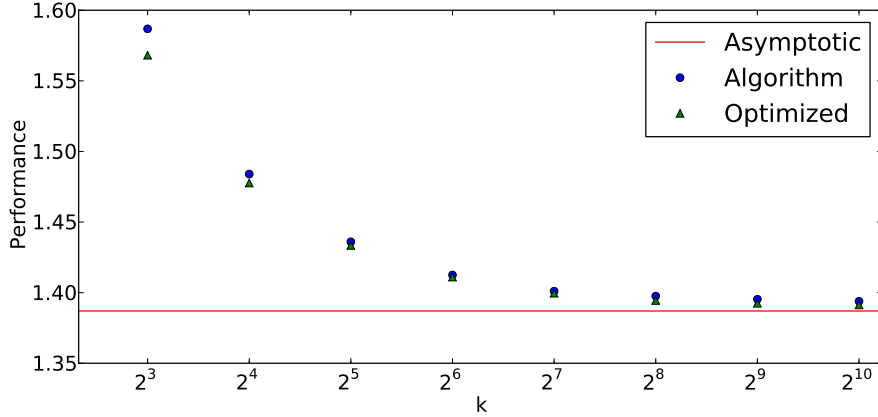


Figure 5: The discrepancy of the algorithm from Sect. 5 for some values of k , compared with the upper bounds for its pattern found via the combinatorial method from Sect. 6. For $k = 1024$, the optimization procedure finds a checkpoint placement with only 1.9% better discrepancy.

Do we find optimal algorithms? One could ask whether the algorithms from Table 1 are optimal, or at least near optimal. There are two steps in above optimization algorithm that prevent this question to be answered positively. First, we are only optimizing over short patterns, and it might be that much larger pattern lengths are necessary for optimal checkpointing algorithms. Second, we do not know how smoothly the optimal discrepancy for fixed pattern P and scaling factor γ behaves with varying γ , i.e., we do not know whether our linear search for γ yields any approximation on the discrepancy λ . However, in experiments we tried all patterns of length $2k$ for $k \in [3, 4, 5]$ and found no better algorithm than for the shorter patterns of length up to k . Moreover, smaller step sizes in the linear search for γ lead only to small improvements, indicating that the discrepancy is continuous in γ . This suggests that the reported algorithms might be near optimal.

7 Existence of Optimal Algorithms

In this section, we prove that optimal algorithms for the checkpointing problem exist, i.e., that there is an algorithm having discrepancy equal to the infimum discrepancy $q^*(k) := \inf_A \text{Perf}(A)$ among all algorithms for k checkpoints.

Theorem 4. *For each k there exists a checkpointing algorithm A for k checkpoints with $\text{Perf}(A) = q^*(k)$, i.e., there is an optimal checkpointing algorithm.*

As we will see throughout this section, this is a non-trivial statement. From the proof of this statement, we gain additional insight in the behavior of good

algorithms. In particular, we show that we can assume without increasing discrepancy that for all i the i -th checkpoint is set by a factor of at least $(1 + 1/k)^{\Theta(i)}$ later than the first checkpoint.

An initial set of checkpoints can be described by a vector $x = (x_1, \dots, x_k)$, $0 \leq x_1 \leq \dots \leq x_k$. Since $x = (0, \dots, 0)$ can never be extended to a checkpointing algorithm of finite discrepancy, we shall always assume $x \neq 0$. Denote by X the set of all initial sets of checkpoints (described by vectors $x \neq 0$ as above), and by X_0 the set of all $x \in X$ with $x_k = 1$.

We say that $A = (t, d)$ is an algorithm for an initial set $x \in X$ of checkpoints if $t_i = x_i$ for all $i \in [k]$. We denote by $q(x) := \inf_A \text{Perf}(A)$, where A runs over all algorithms for x , the *discrepancy* of x . An initial set $x \in X$ is called *optimal* if $q(x) = \inf_{x \in X} q(x) = q^*(k)$.

Lemma 6. *Optimal initial sets of checkpoints exist.*

Proof. Since the discrepancy of an initial set of checkpoints is invariant under scaling, that is, $q(x) = q(\lambda x)$ for all $x \in X$ and $\lambda > 0$, we have $\inf_{x \in X_0} q(x) = \inf_{x \in X_0} q(x)$.

It is not hard to see that $q(\cdot)$ is continuous on X_0 : Let $x, x' \in X_0$ with $|x - x'|_\infty \leq \varepsilon$ and consider an algorithm $A = (t, d)$ for x . We construct an algorithm $A' = (t', d)$ for x' by setting $t'_i = t_i$ for $i > k$. Then $|\text{Perf}(A) - \text{Perf}(A')| \leq 2\varepsilon$, since any interval's length is changed by at most 2ε . This implies $|q(x) - q(x')| \leq 2\varepsilon$ and, thus, shows continuity of $q(\cdot)$.

Now, since $q(\cdot)$ is continuous on X_0 and X_0 is compact, there exists an $x \in X_0$ such that $q(x) = \inf_{x \in X_0} q(x) = q^*(k)$. \square

An easy observation is that if some checkpointing algorithm leads to a vector x of checkpoints at some time, then we may continue from there using any other algorithm for x . The discrepancy of this combined algorithm is at most the maximum of the two discrepancies.

Lemma 7. *Let $A = (t, d)$ be a checkpointing algorithm. Let $i > k$. We call $q_{A,i} = \max_{j \in [k..i]} \bar{\ell}_{t_j}(k+1)/t_j$ the partial discrepancy of A observed in the time up to t_i . Assume that when running A , at time t_i the checkpoints $x = (x_1, \dots, x_k = t_i)$ are active. Let $A' = (t', d')$ be an algorithm for x . Then the checkpointing algorithm obtained from running A until time t_i and then continuing with algorithm A' is a checkpointing algorithm that has discrepancy at most $\max\{q_{A,i}, \text{Perf}(A')\}$. If we run this combined algorithm only until some time t'_j , then the partial discrepancy observed till then is $\max\{q_{A,i}, q_{A',j}\}$.*

Proof. Trivial. \square

The above lemma implies that in the following, we may instead of looking at an arbitrary time simply assume that the algorithm just started, that is, that the current set of checkpoints is the initial one.

The following lemma shows that we can, without loss of discrepancy, assume that an algorithm for the checkpointing problem does not set checkpoints too close together. While also of independent interest, among others because it

shows how to keep additional costs for setting and removing checkpoints low, we shall need this statement in our proof that optimal checkpointing algorithms exist.

Lemma 8. *Let $A = (t, d)$ be an algorithm for the checkpointing problem with $\text{Perf}(A) < k + 1$. Then there is an algorithm $A' = (t', d')$ with the same starting position such that (i) $\text{Perf}(A') \leq \text{Perf}(A)$ and*

$$(ii) \ t'_{k+3} \geq t'_k \left(1 + \frac{\text{Perf}(A)}{k+1 - \text{Perf}(A)} \right) \geq t'_k \left(1 + \frac{1}{k} \right).$$

Proof. Let $r = \text{Perf}(A)/(k+1 - \text{Perf}(A))$ for convenience. By way of contradiction, assume that the lemma is false. Let A be a counter-example such that $i := \min\{i \in \mathbb{N} \mid t_{k+i} \geq 1+r\}$ is minimal (the minimum is well-defined, since for any algorithm the sequence $(t_i)_i$ tends to infinity). Note that $i \geq 4$, since A is a counter-example.

Assume that there is a $j \in [1..i-1]$ such that t_{k+j} in the further run of A is removed (and replaced by the then current time t_x) earlier than both t_{k+j-1} and t_{k+j+1} . Consider the Algorithm A' that arises from A by the following modifications. Let t_y be the checkpoint that was removed to install the checkpoint t_j . Let A' be the checkpointing algorithm that proceeds as A except that t_y is not replaced by t_{k+j} , but by t_x , and t_{k+j} is never created. The only interval which could cause this algorithm to have a worse discrepancy than A is $[t_{k+j-1}, t_{k+j+1}]$. However, this interval contributes $(k+1)(t_{k+j+1} - t_{k+j-1})/t_{k+j+1} \leq (k+1)r/(1+r) \leq \text{Perf}(A)$ to the discrepancy of A' . Hence, $\text{Perf}(A') \leq \text{Perf}(A)$ and A' has fewer checkpoints in the interval $[1, 1+r]$ contradicting the minimality of A . Thus, there is no $j \in [1..i-1]$ such that t_{k+j} is removed earlier than both t_{k+j-1} and t_{k+j+1} (*).

We consider now separately the two cases that t_{k+1} is removed earlier than t_{k+i-2} and vice versa. Note first that $k+1 < k+i-2$ by assumption that $i \geq 4$.

Assume first that t_{k+1} is removed (at some time t_x) earlier than t_{k+i-2} . Then t_k must have been removed even earlier (at some time t_y), otherwise we found a contradiction to (*). Let A' be an algorithm working identically as A , except that at time t_y the checkpoint t_{k+1} is removed (instead of t_k) and at time t_x the checkpoint t_k is removed (instead of t_{k+1}). Since the checkpoint at t_{k+i-2} is still present, the only interval affected by this exchange, namely the one with t_k as left endpoint, has length at most r . Hence as above, this contributes at most $\text{Perf}(A)$ to the discrepancy of A' . The algorithm A' has the property that there is a checkpoint in between t_k and t_{k+i-2} which is removed before these two points. The earliest such checkpoint, call it t_{k+j} , has the property that t_{k+j} is removed earlier than both t_{k+j-1} and t_{k+j+1} , contradicting earlier arguments.

A symmetric argument shows that also t_{k+i-2} being removed before t_{k+1} leads to a contradiction. Consequently, our initial assumption that $i \geq 4$ cannot hold, proving the claim. \square

The following is a global variant of Lemma 8. It shows that any reasonable checkpointing algorithm does not store new checkpoints too often.

Theorem 5. *Let $A = (t, d)$ be a checkpointing algorithm with $\text{Perf}(A) < k - 1$. Then there is an algorithm $A' = (t', d')$ with the same starting position such that (i) $\text{Perf}(A') \leq \text{Perf}(A)$ and (ii) $t'_{i+3} \geq (1 + 1/k) \cdot t'_i$ for all $i \geq k$.*

Proof. Let $j \geq k$ be the smallest index with a small jump, $t_{j+3} < (1 + 1/k)t_j$. Using Lemma 8 (on the remainder of algorithm A starting at time t_j) we can remove this small jump and get an algorithm $A' = (t', d')$ with $\text{Perf}(A') \leq \text{Perf}(A)$ and $t'_{i+3} \geq (1 + 1/k) \cdot t'_i$ for all $k \leq i \leq j$, i.e., we patched the earliest small jump. Iterating this patching procedure infinitely often yields the desired algorithm. \square

Lemma 9. *For any optimal initial set $x = (x_1, \dots, x_k)$, there is an algorithm $A = (t, d)$ such that (i) $q_{A,k+3} = \max_{j \in [k..k+3]} \ell_{t_j}(k+1)/t_j \leq q^*(k)$, (ii) $t_{k+3} \geq t_k(1 + 1/k)$, and the set of checkpoints active at time t_{k+3} is again optimal.*

Proof. By the definition of optimality, for each $n \in \mathbb{N}$ there is an algorithm $A^{(n)}$ for x that has discrepancy at most $q^*(k) + 1/n$. Let $(t_{k+1}^{(n)}, t_{k+2}^{(n)}, t_{k+3}^{(n)})$ denote the corresponding next three checkpoints. By Lemma 8, we may assume that $t_{k+3}^{(n)} \geq t_k(1 + 1/k)$ for all $n \in \mathbb{N}$.

Note that (using the same arguments as in Lemma 5) any algorithm having discrepancy at most 2.5 satisfies $t_{k+i} \leq 6^i t_k$ for any $k \geq 2$. Hence, $(t_{k+1}^{(n)}, t_{k+2}^{(n)}, t_{k+3}^{(n)})_{n \in \mathbb{N}_{\geq 2}}$ is a sequence in the compact space $[t_k, 6^3 t_k]^3$. This sequence has a convergent subsequence with limit $(t_{k+1}, t_{k+2}, t_{k+3})$. Also, since there are only finitely many values possible for $(d_{k+1}^{(n)}, d_{k+2}^{(n)}, d_{k+3}^{(n)})$, this subsequence can be chosen such that this d -tuple is constant, say $(d_{k+1}, d_{k+2}, d_{k+3})$. For this subsequence, also all $k+1$ intervals existing at the three times of interest converge. Consequently, the discrepancy caused by each of them also converges to a value upper bounded by $q^*(k)$. This defines the three steps of algorithm A , satisfying $q_{A,k+3} \leq q^*(k)$.

Similarly, we observe that the set of checkpoints $x^{(n)}$ active at time $t_{k+3}^{(n)}$ when running algorithm $A^{(n)}$ has discrepancy at most $q^*(k) + 1/n$. Consequently, the active checkpoints we get from the limit checkpoints $(t_{k+1}, t_{k+2}, t_{k+3})$ and deletions $(d_{k+1}, d_{k+2}, d_{k+3})$ are again optimal.

Finally, since all $t_{k+3}^{(n)} \geq t_k(1 + 1/k)$, this also holds for t_{k+3} . \square

We are now in position to prove the main result of this section, Theorem 4. For this, we repeatedly apply Lemma 9: We start with an optimal set of checkpoints x . Then we run the algorithm delivered by Lemma 9 for three steps. This creates no partial discrepancy larger than $q^*(k)$ and we end up with another optimal set of checkpoints. From this, we continue to apply Lemma 9 and execute three steps of the algorithm obtained. By Lemma 7, the partial discrepancy of the combined algorithm is again at most $q^*(k)$. Iterating infinitely, this yields an optimal algorithm, which proves Theorem 4.

8 Lower Bound

In this section, we prove a lower bound on the discrepancy of all checkpointing algorithms. For large k we get a lower bound of roughly 1.3, so we have a lower bound that is asymptotically larger than the trivial bound of 1. Moreover, it shows that algorithm BINARY from Sect. 5 is nearly optimal, as for large k the presented lower bound is within 6% of the discrepancy of BINARY.

Theorem 6. *All checkpointing algorithms with k checkpoints have a discrepancy of at least*

$$2 - \ln 2 - O(k^{-1}) \geq 1.306 - O(k^{-1}).$$

The remainder of this section is devoted to the proof of the above theorem. Let $A = (t, d)$ be an arbitrary checkpointing algorithm and let $q' := \text{Perf}(A)$ be its discrepancy. For convenience, we define $q = kq'/(k+1)$ and bound q . Since $q < q'$ this suffices to show a lower bound for the discrepancy of A . For technical reasons we add a *gratis checkpoint* at time t_k that must not be removed by A . That is, even after the removal of the original checkpoint at t_k , there still is the gratis checkpoint active at t_k . Clearly, this can only improve the discrepancy. We analyze the discrepancy of A from time t_k until it deleted $k/(2q)$ of the initial checkpoints¹. More formally, we let t' be the minimal time at which the number of active checkpoints of A contained in $[0, t_k]$ is $k - k/(2q)$. Note that we might have $t' = \infty$, if the checkpointing algorithm A never deletes $k/(2q)$ points from $[0, t_k]$. However, in this case its discrepancy is lower bounded by 1.5.

Lemma 10. *If $t' = \infty$, then $\text{Perf}(A) \geq 1.5$.*

Proof. Consider a large $i > k$ and the algorithm's discrepancy at time t_i . By assumption, there are at most $k - k/(2q)$ active checkpoints in $(t_k, t_i]$. Hence, by comparing with an equidistant spread we can bound the discrepancy (at time t_i) by

$$\text{Perf}(A) \geq \frac{k+1}{t_i} \cdot \frac{t_i - t_k}{k(1 - 1/(2q))} \geq \frac{2q}{2q-1} \left(1 - \frac{t_k}{t_i}\right).$$

Letting $i \rightarrow \infty$, so that $t_i \rightarrow \infty$, we obtain

$$\text{Perf}(A) \geq \frac{2q}{2q-1} \geq \frac{2 \text{Perf}(A)}{2 \text{Perf}(A) - 1},$$

(by definition of q and $x \mapsto \frac{2x}{2x-1}$ being monotonically decreasing). This inequality solves to the desired $\text{Perf}(A) \geq 1.5$. \square

Hence, in the following we can assume that $t' < \infty$. We partition the intervals that exist at time t' into three types:

¹To be precise we should round $\frac{k}{2q}$ to one of its nearest integers. When doing so, all calculations in the remainder of this section go through as they are; this only slightly increases the hidden constant in the error term $O(k^{-1})$.

1. Intervals existing both at time t_k and t' . These intervals are contained in $[0, t_k]$.
2. Intervals that are contained in $[0, t_k]$, but did not exist at time t_k . These intervals were created by the removal of some checkpoint in $[0, t_k]$ after time t_k .
3. Intervals contained in $[t_k, t']$.

Note that we need the gratis checkpoint at t_k in order for these definitions to make sense, as otherwise there could be an interval overlapping t_k .

Let \mathcal{L}_i denote the set of intervals of type i for $i \in \{1, 2, 3\}$, and set $k_i := |\mathcal{L}_i|$. Let $\mathcal{L}_2 = \{I_1, \dots, I_{k_2}\}$, where the intervals are ordered by their creation times $\tau_1 \leq \dots \leq \tau_{k_2}$. Since each interval in \mathcal{L}_2 contains at least one deleted point we have

$$k_2 \leq \frac{k}{2q},$$

and we set $m := \frac{k}{2q} - k_2$. Then m counts the number of deleted checkpoints in $[0, t_k]$ that did not create an interval in \mathcal{L}_2 , but some strict sub-interval of an interval in \mathcal{L}_2 . We call these m removed checkpoints *free*.

We first bound the length of the intervals in \mathcal{L}_1 and \mathcal{L}_2 .

Lemma 11. *The length of any interval in \mathcal{L}_1 is at most qt_k/k .*

Proof. As all intervals in \mathcal{L}_1 already are present at time t_k and the algorithm has discrepancy q' , we have for any $I \in \mathcal{L}_1$

$$(k+1)|I|/t_k \leq q' = (k+1)q/k.$$

The bound follows. □

Lemma 12. *The length of any interval $I_i \in \mathcal{L}_2$ is at most*

$$|I_i| \leq \frac{t_k}{k/q - m - i}.$$

Proof. As the algorithm has discrepancy q' , we know

$$|I_i| \leq q\tau_i/k. \tag{6}$$

In the following we bound τ_i , the time of creation of I_i . At time τ_i there are at most $m+i$ intervals in \mathcal{L}_3 , since at most m free checkpoints and i checkpoints from the creation of I_1, \dots, I_i are available. Comparing with an equidistant spread of $m+i$ checkpoints in $[t_k, \tau_i]$ and the algorithm's discrepancy, the longest interval L in $[t_k, \tau_i]$ (at time τ_i) has length

$$\frac{\tau_i - t_k}{m+i} \leq |L| \leq \frac{q\tau_i}{k}.$$

Rearranging the outer inequality yields a bound on τ_i of

$$\tau_i \leq \frac{kt_k}{k - (m + i)q}.$$

Substituting this into (6) yields the desired result. \square

Furthermore, we need a relation between k_1, k, m , and q .

Lemma 13. *We have*

$$k_1 = k + m - k/q + 1.$$

Proof. As the intervals in \mathcal{L}_1 and \mathcal{L}_2 partition $[0, t_k]$, there are $k_1 + k_2$ intervals left in $[0, t_k]$ at time t' . Note that each but one such interval has its left endpoint among the k active checkpoints from time t_k (the one exception having as left endpoint 0). Hence, there are $k_1 + k_2 - 1$ checkpoints left in $[0, t_k]$. Comparing with the number $k_2 + m$ of deleted checkpoints in $[0, t_k]$ until time t' and their overall number k yields

$$(k_2 + m) + (k_1 + k_2 - 1) = k.$$

Rearranging this and plugging in $k_2 = \frac{k}{2q} - m$ (which holds by definition of m) yields the desired result. \square

Now we use our bounds on the length of intervals from \mathcal{L}_1 and \mathcal{L}_2 to find a bound on q . Note that the intervals in \mathcal{L}_1 and \mathcal{L}_2 partition $[0, t_k]$, so that

$$t_k = \sum_{I \in \mathcal{L}_1} |I| + \sum_{I' \in \mathcal{L}_2} |I'|.$$

Using Lemmas 11 and 12, we obtain

$$t_k \leq k_1 \frac{qt_k}{k} + \sum_{i=1}^{k_2} \frac{t_k}{k/q - m - i}.$$

Substituting k_1 using Lemma 13 yields

$$\begin{aligned} t_k &\leq [k + m - k/q + 1] qt_k/k + \sum_{i=1}^{k/(2q)-m} \frac{t_k}{k/q - m - i} \\ &= t_k \left(q - 1 + m \frac{q}{k} + O(k^{-1}) + \sum_{i=1}^{k/(2q)-m} \frac{1}{k/q - m - i} \right). \end{aligned} \quad (7)$$

Recall that $H_n = \sum_{1 \leq i \leq n} i^{-1}$ is the n -th harmonic number. Rearranging (7) yields

$$q \geq 2 - m \frac{q}{k} - O(k^{-1}) - H_{k/q-m-1} + H_{k/(2q)-1}.$$

Observe that we have $m_k^q + H_{k/q-m-1} \leq H_{k/q-1}$, implying

$$\begin{aligned} q &\geq 2 + H_{k/(2q)-1} - H_{k/q-1} - O(k^{-1}) \\ &\geq 2 + H_{k/(2q)} - H_{k/q} - O(k^{-1}), \end{aligned}$$

since we can hide the last summands of $H_{k/(2q)}$ and $H_{k/q}$ by $O(k^{-1})$. In combination with the asymptotic behavior of $H_n = \ln n + \gamma + O(n^{-1})$, where γ is the Euler-Mascheroni constant, we obtain

$$\begin{aligned} q &\geq 2 + \ln(k/(2q)) - \ln(k/q) - O(k^{-1}) \\ &= 2 - \ln(2) - O(k^{-1}). \end{aligned}$$

This finishes the proof of Theorem 6.

References

- [1] Lauri Ahlroth, Olli Potttonen, and André Schumacher. Approximately uniform online checkpointing with bounded memory. *Algorithmica*, 2013. To appear.
- [2] Marshall W. Bern, Daniel H. Greene, Arvind Raghunathan, and Madhu Sudan. On-line algorithms for locating checkpoints. *Algorithmica*, 11(1):33–52, 1994.
- [3] Kaniyanthra Mani Chandy and Chittoor V. Ramamoorthy. Rollback and recovery strategies for computer programs. *IEEE Transactions on Computers*, C-21:546–556, 1972.
- [4] E. N. (Mootaz) Elnozahy, Lorenzo Alvisi, Yi-Min Wang, and David B. Johnson. A survey of rollback-recovery protocols in message-passing systems. *ACM Computing Surveys*, 34(3):375–408, 2002.
- [5] Erol Gelenbe. On the optimum checkpoint interval. *Journal of the ACM*, 26(2):259–270, 1979.
- [6] Vincent Heuveline and Andrea Walther. Online checkpointing for parallel adjoint computation in PDEs: Application to goal-oriented adaptivity and flow control. In Wolfgang E. Nagel, Wolfgang V. Walter, and Wolfgang Lehner, editors, *Euro-Par 2006 Parallel Processing*, volume 4128 of *Lecture Notes in Computer Science*, pages 689–699. Springer-Verlag, 2006.
- [7] Fredrik Österlind, Adam Dunkels, Thiemo Voigt, Nicolas Tsiftes, Joakim Eriksson, and Niclas Finne. Sensornet checkpointing: Enabling repeatability in testbeds and realism in simulations. In Utz Roedig and Cormac J. Sreenan, editors, *Wireless Sensor Networks*, volume 5432 of *Lecture Notes in Computer Science*, pages 343–357. Springer-Verlag, 2009.

- [8] P. Stumm and A. Walther. New algorithms for optimal online checkpointing. *SIAM Journal on Scientific Computing*, 32(2):836–854, 2010.
- [9] S. Toueg and Ö. Babaoglu. On the optimum checkpoint selection problem. *SIAM Journal on Computing*, 13(3):630–649, 1984.
- [10] Sangho Yi, D. Kondo, and A. Andrzejak. Reducing costs of spot instances via checkpointing in the Amazon elastic compute cloud. In *IEEE 3rd International Conference on Cloud Computing (CLOUD 2010)*, pages 236–243, 2010.